Parametric Inference for Diffusion Processes with Noisy, Nonsynchronous Observations

Teppei Ogihara

The Institute of Statistical Mathematics

Mar. 18. 2015. SAPS X

Outline

- In this talk, we study maximum-likelihood-type estimation, Bayes-type estimation and their asymptotics for parametric diffusion processes with high-frequency data
- In particular, we focus on two specific problems on analysis of high-frequncy data:
  - the presence of market microstructure noise
  - nonsynchronous observations
- We consider maximum-likelihood-type estimation of parameters under these problems.

Introduction

- Volatility estimation under the presence of market microstructure noise is studied—for example, in Zhang, Mykland, and Aït-Sahalia (2005), Barndorff-Nielsen et al. (2008), and Podolskij and Vetter (2009)—by using various data-averaging or resampling methods to reduce the influence of noise
- Covariation estimation under nonsynchronous observations are independently studied by Hayashi and Yoshida (2005,2008,2011) and Malliavin and Mancino (2002,2009).
- There are also studies of covariation estimation under the simultaneous presence of microstructure noise and nonsynchronous observations:
  - Barndorff-Nielsen et al. (2011) — a kernel based method
  - Christensen, Kinnebrock, and Podolskij (2010)
    — a modulated realized covariance and a pre-averaged HY
  - Aït-Sahalia, Fan, and Xiu (2010)
    — a method using MLE with constant diffusion coefficients
  - Bibinger et al. (2014) — a local method of moments

Introduction

- While the above studies concern estimators under non- or semi-parametric settings, there are also studies about parametric inference of diffusion processes with high-frequency observations:
  - Genon-Catalot and Jacod (1994) constructed quasi-likelihood function and studied an estimator that maximizes it for a model of equi-distant observations of diffusion processes without noise
  - Gloter and Jacod (2001b) studied an estimator based on a quasi-likelihood function with noisy observations
  - Ogihara and Yoshida (2014) studied a maximum-likelihood-type estimator and a Bayes-type estimator on nonsynchronous observations without noise
- The theory of random fields of (quasi-)likelihood ratios, initiated by Ibragimov and Has'minskii and developed by Kutoyants and Yoshida, enables us to deduce asymptotic properties of ML- and Bayes-type estimators

Introduction

- One advantage of ML- and Bayes-type estimators is that they are asymptotically efficient in many models
- If a statistical model has the local asymptotic mixed normality (LAMN) property, then the results in Jeganathan (1982,1983) give asymptotic minimal variance of estimators. Estimators which attain this bound are called asymptotically efficient
- For above models, Gobet (2001), Gloter and Jacod (2001a), and Ogihara (2014) proved the LAMN(LAN) properties of the corresponding statistical models and asymptotic efficiency of maximum-likelihood-type estimators are obtained
  - Gloter and Jacod (2001a) concerns the LAN property for models with deterministic diffusion coefficients. The LAMN property for models of general diffusion processes seems to be unsolved

Introduction

- Bibinger et al.(2014) showed a lower bound of asymptotic variance of estimators in semi-parametric Cramér-Rao sense
- We need the LAN or LAMN property of the statistical model to obtain asymptotic efficiency of a parametric model. To the best of my knowledge, this has not been studied for statistical models of noisy, nonsynchronous observations
- In this talk, we examines consistency and asymptotic mixed normality of a ML-type estimator and a Bayes-type estimator based on a quasi-likelihood function with noisy, nonsynchronous observations
- We also study the LAN property of this model when diffusion coefficients are constants, as well as the asymptotic efficiency of our estimators

Setting

Processes Let $X = \{X_t\}_{0 \le t \le T}$ be a two-dimensional stochastic process satisfying the following SDE:

$$dX_t = \mu_t dt + b(t, X_t, \sigma_*) dW_t, \quad t \in [0, T]. \tag{1}$$

$\{W_t\}_{0 \le t \le T}$: a two-dimensional standard Brownian motion,
b : a 2×2 matrix-valued function (known),
μ_t: a two-dimensional locally bounded process (unknown),
$\sigma_* \in \mathbb{R}^d$: an unknown parameter

Nonsynchronous Observations $\{S_i^{n,p}\}_{i=0}^{\ell_{p,n}} \subset [0, T]$: observation times of X^p (random), $0 = S_0^{n,p} < S_1^{n,p} < \cdots < S_{\ell_{p,n}}^{n,p} = T$,

$$r_n = \max_{i,p}(S_i^{n,p} - S_{i-1}^{n,p}) \to^p 0 \quad \text{as} \quad n \to \infty \quad : \text{a high-frequency limit}$$

Observation noise $\{\epsilon_i^{n,p}\}_{i \in \mathbb{Z}_+}$: i.i.d., $E[\epsilon_0^{n,p}] = 0$, $E[(\epsilon_0^{n,p})^2] = v_{p,*}$, $E[(\epsilon_0^{n,p})^q] < \infty$ $(q > 0)$, $v_{p,*}$: unknown

Our Purpose To estimate the unknown parameter σ_* with noisy, nonsynchronous observations $Y_i^p = X_{S_i^{n,p}}^p + \epsilon_i^{n,p}$

Construction of the MLE

- Let $\{b_n\}_n$ and $\{l_n\}_n$ be sequences of positive numbers such that $0 < \mathrm{P}\text{-}\lim_{n\to\infty}(\#\{S_i^{n,p}\}_i/b_n) < \infty$ a.s., $l_n \in \mathbb{N}$, $l_n \to \infty$, $\exists\epsilon > 0$, $l_n b_n^{-1/2-\epsilon} \to 0$, $l_n^{-1} b_n^{4/13+\epsilon} \to 0$ and $s_m = T l_n^{-1} m$ $(0 \leq m \leq l_n)$. We divide observations into disjoint intervals $\{[s_{m-1}, s_m)\}_{m=1}^{l_n}$ and construct a quasi-likelihood function
- Let $\Delta Y_i^p = Y_i^p - Y_{i-1}^p$. Then

$$\begin{aligned} E_m[\Delta Y_{i_1}^p \Delta Y_{i_2}^p] &= E_m\left[\prod_{k=1}^2 \left(\int_{I_{i_k}^p} (\mu_t dt + b dW_t) + \epsilon_{i_k}^{n,p} - \epsilon_{i_k-1}^{n,p}\right)\right] \\ &\sim |b_m^p(\sigma_*)|^2 |I_{i_1}^p| \delta_{i_1 i_2} + v_{p,*}(M_{p,m})_{i_1 i_2}, \\ E_m[\Delta Y_{i_1}^1 \Delta Y_{i_2}^2] &\sim b_m^1 \cdot b_m^2(\sigma_*) |I_{i_1}^1 \cap I_{i_2}^2|, \end{aligned}$$

where $I_i^p = [S_{i-1}^{n,p}, S_i^{n,p})$, $M_{p,m} = \{2\delta_{ij} - \delta_{|i-j|=1}\}_{i,j}$, $E_m[\cdot] = E[\cdot|\mathcal{F}_{s_{m-1}}]$, k_m^p is observation counts of Y^p in $[s_{m-1}, s_m)$, and $b_m^p(\sigma) = b^p(s_{m-1}, ((k_{m-1}^\alpha)^{-1} \sum_{i; S_i^{n,\alpha} \subset [s_{m-2}, s_{m-1})} Y_i^\alpha)_\alpha, \sigma)$

Construction of the MLE

- Therefore a quasi-likelihood function $H_n(\sigma, v)$ is defined by

$$H_n(\sigma, v) = -\frac{1}{2}\sum_{m=2}^{l_n}(Z_m^\top S_m^{-1}(\sigma, v)Z_m + \log\det S_m(\sigma, v)),$$

where $Z_m = ((\Delta Y_i^1)_{i;I_i^1 \subset [s_{m-1}, s_m)}, (\Delta Y_j^2)_{j;I_j^2 \subset [s_{m-1}, s_m)})$ and

$$\begin{aligned}
&S_m(\sigma, v) \\
= &\begin{pmatrix} |b_m^1|^2(\sigma)\mathrm{diag}((|I_i^1|)_i) & b_m^1 \cdot b_m^2(\sigma)\{|I_i^1 \cap I_j^2|\}_{i,j} \\ b_m^1 \cdot b_m^2(\sigma)\{|I_i^1 \cap I_j^2|\}_{j,i} & |b_m^2|^2(\sigma)\mathrm{diag}((|I_j^2|)_j) \end{pmatrix} \\
&+ \begin{pmatrix} v_1 M_{1,m} & 0 \\ 0 & v_2 M_{2,m} \end{pmatrix} \qquad (v = (v_1, v_2))
\end{aligned}$$

- If $\epsilon_i^{n,k}$ is non-Gaussian, the *true* log-likelihood function is not approximated by H_n. But we can see that such misspecification of the noise distribution do not affect estimation of σ

Construction of the MLE

- Let $\{\hat{v}_n\}_{n\in\mathbb{N}}$ be an estimator of the unknown parameter $v_* = (v_{1,*}, v_{2,*})$, where $\{b_n^{1/2}(\hat{v}_n - v_*)\}_n$ is tight. Then a ML-type estimator $\hat{\sigma}_n$ of σ_* is defined by

$$\hat{\sigma}_n = \mathrm{argmax}_\sigma H_n(\sigma, \hat{v}_n)$$

- For example,

$$\hat{v}_{p,n} = (2\ell_{p,n})^{-1} \sum_{i=1}^{\ell_{p,n}} (\Delta Y_i^p)^2$$

  satisfies the above condition

- We can use the (adaptive or simultaneous) ML-type estimator of v for $\hat{v}_n$ if $\epsilon_i^{n,k}$ is Gaussian. But the above nonparametric estimator of v allows us to use non-Gaussian noise. The choice of estimators for v does not affect efficiency of $\hat{\sigma}_n$.

Main results

Let $\underline{r}_n = \min_{p,i}(S_i^{n,p} - S_{i-1}^{n,p})$. We assume the following:

$A1.$ μ_t is locally bounded, $b(t,x,\sigma)$ is smooth w.r.t. (t,x,σ) and $bb^\top$ is positive definite for any (t,x,σ)

$A2.$ $\inf_{\sigma_1 \neq \sigma_2}(|bb^\top(t,x,\sigma_1) - bb^\top(t,x,\sigma_2)|/|\sigma_1 - \sigma_2|) > 0$ for any (t,x)

$A3.$ $\forall \delta > 0$, $r_n = O_p(b_n^{-1+\delta})$ and $\underline{r}_n^{-1} = O_p(b_n^{1+\delta})$

$A4.$ $\exists \eta \in (0,1/2)$, $\exists \{a_0^p(t)\}_{0 \leq t \leq T, p=1,2}$: positive-valued stochastic processes s.t. $a_0^p(t)$ is C^1 w.r.t. t and

$$l_n^{-1} b_n^{1/2} \max_{1 \leq l \leq L_n} \left| b_n^{-1}(s''_{n,l} - s'_{n,l})^{-1} \sum_{i; I_i^p \subset [s'_{n,l}, s''_{n,l})} 1 - a_0^p(s'_{n,l}) \right| \to^p 0$$

as $n \to \infty$, where $\{[s'_{n,l}, s''_{n,l})\}_{n,l} \subset [0,T]$ are arbitrary disjoint intervals satisfying

$$0 < \inf_{n,l}(b_n^{1-\eta}(s''_{n,l} - s'_{n,l})) \leq \sup_{n,l}(b_n^{1-\eta}(s''_{n,l} - s'_{n,l})) < \infty$$

※ Intuitively, $[A4]$ shows the law of large numbers in any local intervals

Main results

Theorem 1

Assume $[A1]$-$[A4]$*. Then* $\Gamma := P\text{-}\lim_{n\to\infty}(-b_n^{-1/2}\partial_\sigma^2 H_n(\sigma_*, v_*))$ *is positive definite a.s. and* $\exists\zeta$ *:* d*-dimensional standard normal random variable independent of* Γ *such that*

$$b_n^{1/4}(\hat{\sigma}_n - \sigma_*) \to^{s\text{-}\mathcal{L}} \Gamma^{-1/2}\zeta \quad \text{and} \quad -b_n^{-1/2}\partial_\sigma^2 H_n(\hat{\sigma}_n, \hat{v}_n) \to^p \Gamma.$$

In particular, $(-\partial_\sigma^2 H_n(\hat{\sigma}_n, \hat{v}_n))^{1/2}(\hat{\sigma}_n - \sigma_*) \to^d N(0, I_d)$.

On asymptotic efficiency of the estimator

Theorem 2

Assume $[A1]$-$[A4]$ and that $\mu_t = \mu(t, X_t)$, $\epsilon^{n,p}$ is Gaussian and $b(t, x, \sigma)$ does not depend on (t, x). Then the LAN property for our model with noisy, nonsynchronous observations holds and $\hat{\sigma}_n$ is asymptotically efficient

Convergence of moments and Bayes-type estimator

- Let $\pi(\sigma)$ be a prior density function, which is continuous, bounded and $\inf_\sigma \pi(\sigma) > 0$. A Bayes-type estimator $\tilde{\sigma}_n$ for the quadratic loss function is defined as

$$\tilde{\sigma}_n = \left(\int \exp(H_n(\sigma, \hat{v}_n))\pi(\sigma)d\sigma \right)^{-1} \int \sigma \exp(H_n(\sigma, \hat{v}_n))\pi(\sigma)d\sigma$$

Theorem 3

Assume some stronger conditions than $[A1]$-$[A4]$ (some conditions on moments of μ_t, $b(t, X_t, \sigma)$ and the convergent sequence in $[A4]$, etc). Then

$$\begin{aligned} E[\mathbf{Y} f(b_n^{1/4}(\hat{\sigma}_n - \sigma_*))] &\to E[\mathbf{Y} f(\Gamma^{-1/2}\zeta)], \\ E[\mathbf{Y} f(b_n^{1/4}(\tilde{\sigma}_n - \sigma_*))] &\to E[\mathbf{Y} f(\Gamma^{-1/2}\zeta)] \end{aligned}$$

for any bounded random variable $\mathbf{Y}$ on $(\Omega, \mathcal{F})$ and any continuous function f of polynomial growth, where Γ and ζ are the ones in Theorem 1.

On a strong identifiability condition

- To obtain Theorem 3, we need a strong identifiability condition : for any $L > 0$, there exists $C_L > 0$ such that

$$P[\inf_{\sigma \neq \sigma_*}((-\mathcal{Y}(\sigma))/|\sigma - \sigma_*|^2) < 1/r] < C_L/r^L \ (r > 0),$$

where $\mathcal{Y}(\sigma) = P\text{-}\lim_{n\to\infty} b_n^{-1/2}(H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n))$.
- This condition is non-trivial for statistical models with non-ergodic limit. Uchida and Yoshida (2013) studied sufficient conditions of this condition for a statistical model of diffusion processes with equi-distant observation without noise
- We can prove that the strong identifiability condition for our model holds if the corresponding condition for equi-distant observation without noise is satisfied. So we can apply the results of Uchida and Yoshida (2013)
- One simple example of sufficient conditions is $\inf_{\sigma_1 \neq \sigma_2, t, x}(|bb^\top(t, x, \sigma_1) - bb^\top(t, x, \sigma_2)|/|\sigma_1 - \sigma_2|) > 0$

Simulation

- For the simple model :

$$\begin{cases} dX_t^1 &= \sigma_1 dW_t^1 \\ dX_t^2 &= \sigma_3 dW_t^1 + \sigma_2 dW_t^2 \end{cases}$$

  we calculate the MLE $\hat{\sigma}_n$, where $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ and $Y_0 = 0$
- Let $\{\bar{N}_t^1\}, \{\bar{N}_t^2\}$ be independent Poisson processes with parameter λ_1, λ_2 and observations are given by $S_i^{n,p} = \inf\{t \geq 0; \bar{N}_{nt}^p \geq i\}$
- $Y_i^p = X_{S_i^{n,p}}^p + \epsilon_i^{n,p}$: observations, where $\{\epsilon_i^{n,p}\}_i \sim^{i.i.d.} N(0, v_{p,*})$
- Let $b_n \equiv n$, $l_n \equiv [n^{3/8}]$
- Let $\hat{v}_{p,n} = \sum_i (\Delta Y_i^p)^2/(2\ell_{p,n})$ and $\hat{\sigma}_n$ be the MLE calculated by it
- We also consider a plug-in estimator $\hat{v}'_{p,n} = \hat{v}_{p,n} - |b^p(\hat{\sigma}_n)|^2 T/(2\ell_{p,n})$ and corresponding MLE $\hat{\sigma}'_n$. Moreover, let $\hat{\sigma}''_n$ be MLE obtained by setting $\hat{v}_n \equiv v_*$

Simulation

Table: Simulation results of MLE : $T = 1, (\lambda_1, \lambda_2) = (1, 1)$, $(\sigma_{1,*}, \sigma_{2,*}, \sigma_{3,*}) = (1, \sqrt{1 - 0.5^2}, 0.5), v_* = (0.001, 0.001)$

		σ_1	σ_2	σ_3	v_1	v_2
	true values	1	0.866	0.5	0.001	0.001
$n = 1000$	$\hat{\sigma}_n$	0.899	0.783	0.456	0.00150	0.00151
		(0.041)	(0.041)	(0.058)	(0.00008)	(0.00007)
	$\hat{\sigma}'_n$	0.972	0.849	0.493	0.00109	0.00110
		(0.047)	(0.047)	(0.061)	(0.00007)	(0.00007)
	$\hat{\sigma}''_n$	0.999	0.873	0.507	-	-
		(0.046)	(0.046)	(0.061)	-	-
$n = 5000$	$\hat{\sigma}_n$	0.967	0.831	0.484	0.00110	0.00110
		(0.027)	(0.030)	(0.042)	(0.00003)	(0.00003)
	$\hat{\sigma}'_n$	1.000	0.859	0.499	0.00101	0.00101
		(0.029)	(0.032)	(0.044)	(0.00003)	(0.00003)
	$\hat{\sigma}''_n$	1.004	0.862	0.501	-	-
		(0.029)	(0.030)	(0.044)	-	-

- $\hat{v}_{p,n}$ (the simple estimator) has upper bias and $\hat{\sigma}_n$ has lower bias
- The performance of $\hat{\sigma}''_n$ and $\hat{\sigma}'_n$ are similar

Simulation

Table: Simulation results of MLE : $T = 1, (\lambda_1, \lambda_2) = (1, 1)$, $(\sigma_{1,*}, \sigma_{2,*}, \sigma_{3,*}) = (1, \sqrt{1 - 0.5^2}, 0.5), v_* = (0.005, 0.005)$

		σ_1	σ_2	σ_3	v_1	v_2
	true values	1	0.866	0.5	0.005	0.005
$n = 1000$	$\hat{\sigma}_n$	0.955	0.834	0.472	0.00546	0.00551
		(0.064)	(0.061)	(0.090)	(0.00032)	(0.00032)
	$\hat{\sigma}'_n$	0.990	0.865	0.488	0.00500	0.00505
		(0.069)	(0.066)	(0.095)	(0.00032)	(0.00033)
	$\hat{\sigma}''_n$	0.991	0.868	0.490	-	-
		(0.070)	(0.067)	(0.094)	-	-
$n = 5000$	$\hat{\sigma}_n$	0.983	0.855	0.488	0.00509	0.00511
		(0.048)	(0.038)	(0.059)	(0.00012)	(0.00011)
	$\hat{\sigma}'_n$	0.992	0.863	0.493	0.00500	0.00501
		(0.049)	(0.038)	(0.060)	(0.00012)	(0.00011)
	$\hat{\sigma}''_n$	0.992	0.863	0.493	-	-
		(0.048)	(0.038)	(0.059)	-	-

Simulation

- We can also consider an estimator $\hat{\sigma}_{1,n}\hat{\sigma}_{3,n}T$ of the quadratic covariation $\langle X^1, X^2\rangle_T$ and it is also efficient in this model
- We compare the performance of this estimator with existing non(semi)parametric estimators
- We use pre-averaged HY (PHY) and Modulated Realized Covariance (MRC) by Christensen, Kinnebrock and Podolskij (2010), local method of moments (LMM) by Bibinger et al. (2014) and Generalized Multiscale Estimator(GME) by Bibinger (2011) as comparing estimators

※ These estimators can be calculated by using "cce" function of "yuima" package in R, except LMM

Simulation

Table: Comparison between estimators of $\langle X^1, X^2 \rangle_T$: $T = 1, (\lambda_1, \lambda_2) = (1, 1)$, $(\sigma_{1,*}, \sigma_{2,*}, \sigma_{3,*}) = (1, \sqrt{1 - 0.5^2}, 0.5)$, $\langle X^1, X^2 \rangle_T = \sigma_{1,*}\sigma_{3,*}T = 0.5$

	n	MLE	PHY	GME	MRC	MRC2	LMM
$v_1 = 0.001$	500	0.458	0.522	0.511	0.538	0.522	0.391
$v_2 = 0.001$		(0.078)	(0.143)	(0.142)	(0.257)	(0.129)	(0.079)
	1000	0.480	0.502	0.490	0.485	0.502	0.468
		(0.070)	(0.133)	(0.129)	(0.180)	(0.118)	(0.076)
	5000	0.500	0.504	0.503	0.502	0.506	0.505
		(0.049)	(0.088)	(0.077)	(0.131)	(0.079)	(0.073)
$v_1 = 0.005$	500	0.499	0.499	0.497	0.524	0.528	0.442
$v_2 = 0.005$		(0.137)	(0.176)	(0.190)	(0.242)	(0.175)	(0.126)
	1000	0.485	0.478	0.466	0.468	0.483	0.507
		(0.108)	(0.163)	(0.137)	(0.192)	(0.122)	(0.118)
	5000	0.490	0.488	0.489	0.505	0.492	0.506
		(0.072)	(0.103)	(0.095)	(0.125)	(0.086)	(0.081)

※ We used $\hat{\sigma}'_n$ as MLE. We used the default of the "cce" function for parameters of other estimators, except $\theta = 1/3$ for MRC2 ($\theta = 0.15$ for PHY, $\theta = 1$ for MRC, $J = 30$, $h^{-1} = 10$ for LMM)

Concluding Remarks

Summary

- Construction of a ML-type estimator and a Bayes type estimator for the model with noisy, nonsynchronous observations
- Asymptotic mixed normality and asymptotic efficiency (if the latent process is a BM) of ML-type estimator
- Results on a Bayes-type estimator and convergence of moments
- Better performance of the estimator compared with other estimators of the quadratic covariation

References

1. Aït-Sahalia, Y., Mykland, P.A., Zhang, L. (2005): How often to sample a continuous-time process in the presence of market microstructure noise, The Review of Financial Studies, 18, 351-416.
2. Barndorff-Nielsen, O. E., Hansen, P. R., Lude, A., Shephard, N. (2008): Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise, Econometrica, 76 (6), 1481-1536.
3. Genon-Catalot, V., Jacod, J. (1993): On the estimation of the diffusion coefficient for multi-dimensional diffusion processes, Annals of Institute of Henri Poincare, 29, 119-151.
4. Gloter, A., Jacod, J. (2001a): Diffusions with measurement errors. I. Local asymptotic normality, ESAIM: Probability and Statistics, 5, 225-242.
5. Gloter, A., Jacod, J. (2001b): Diffusions with measurement errors. II. Optimal estimators, ESAIM: Probability and Statistics, 5, 243-260.
6. Gobet, E. (2001): Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach, Bernoulli, 7, 899-912.

References

7 Hayashi, T., Yoshida, N. (2005): On covariance estimation of non-synchronously observed diffusion processes, Bernoulli 11 (2), 359-379.

8 Jeganathan, P. (1982): On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal, Sankhya, 44, Series A, 173-212.

9 Jeganathan, P. (1983): Some asymptotic properties of risk functions when the limit of the experiment is mixed normal, Sankhya, 45, Series A, 66-87.

10 Ogihara, T. (2014): Local asymptotic mixed normality property for nonsynchronously observed diffusion processes, Bernoulli, preprint.

11 Ogihara, T. and Yoshida, N. (2014): Quasi-likelihood analysis for nonsynchronously observed diffusion processes, Stochastic Processes and their Applications, 124, 2954-3008.

12 Uchida, M., Yoshida, N. (2013): Quasi likelihood analysis for volatility and nondegeneracy of statistical random field, Stochastic Processes and their Applications, 123, 2851-2876.

13 Zhang, L., Mykland, P.A., Aït-Sahalia, Y. (2005): A tale of two time scales: determining integrated volatility with noisy high-frequency data. Journal of American Statistical Association, 100(472), 1394-1411.